

# Flagging and Correction of Pattern Noise in the *Kepler* Focal Plane Array

Jeffery J. Kolodziejczak<sup>\*a</sup>, Douglas A. Caldwell<sup>b</sup>, Jeffrey E. Van Cleve<sup>b</sup>, Bruce D. Clarke<sup>b</sup>,  
Jon M. Jenkins<sup>b</sup>, Miles T. Cote<sup>b</sup>, Todd C. Klaus<sup>b</sup>, Vic S. Argabright<sup>c</sup>

<sup>a</sup>Space Science Office, VP62, NASA Marshall Space Flight Center, Huntsville AL USA 35812;

<sup>b</sup>SETI Institute/NASA Ames Research Center, MS 244-30, Moffett Field, CA USA 94035;

<sup>c</sup>Ball Aerospace & Technologies Corp., 1600 Commerce Street, Boulder, CO USA 80301;

<sup>d</sup>Orbital Sciences Corporation/NASA Ames Research Center, M/S 244-30, Moffett Field, CA, USA 94305;

<sup>e</sup>NASA Ames Research Center, M/S 244-30, Moffett Field, CA, USA 94305

## ABSTRACT

In order for *Kepler* to achieve its required <20 PPM photometric precision for magnitude 12 and brighter stars, instrument-induced variations in the CCD readout bias pattern (our "2D black image"), which are either fixed or slowly varying in time, must be identified and the corresponding pixels either corrected or removed from further data processing. The two principle sources of these readout bias variations are crosstalk between the 84 science CCDs and the 4 fine guidance sensor (FGS) CCDs and a high frequency amplifier oscillation on <40% of the CCD readout channels. The crosstalk produces a synchronous pattern in the 2D black image with time-variation observed in <10% of individual pixel bias histories. We will describe a method of removing the crosstalk signal using continuously-collected data from masked and over-clocked image regions (our "collateral data"), and occasionally-collected full-frame images and reverse-clocked readout signals. We use this same set to detect regions affected by the oscillating amplifiers. The oscillations manifest as time-varying moiré pattern and rolling bands in the affected channels. Because this effect reduces the performance in only a small fraction of the array at any given time, we have developed an approach for flagging suspect data. The flags will provide the necessary means to resolve any potential ambiguity between instrument-induced variations and real photometric variations in a target time series. We will also evaluate the effectiveness of these techniques using flight data from background and selected target pixels.

**Keywords:** *Kepler*, photometer, CCD, noise, pattern noise, CCD readout, crosstalk, image analysis, photometry

## 1. INTRODUCTION

Launched on March 6, 2009, the 3 1/2 year *Kepler* Mission<sup>1</sup> is to detect transiting exoplanets and determine the frequency of Earth-size planets in the habitable zones of sun-like stars. Previously, authors have described mission design and overall performance<sup>2</sup>, photometer design<sup>3,4</sup>, in-flight instrument performance<sup>5,8</sup>, and the overall data processing scheme<sup>6</sup>. Caldwell et al.<sup>5</sup> describes several nonstationary image artifacts that are present in *Kepler* data and discusses their impact on photometric precision. This paper describes a method for recognizing and removing or flagging some of these artifacts. Specifically, the content includes a derivation of the need to develop this capability, a description of the methods employed and a summary of the results from applying it to a short period of flight data. Prior to this we provide a brief introduction to the instrument and data modes.

The *Kepler* focal plane consists of 84 separate science readout channels (identified as module#.output#) and four fine guidance sensor (FGS) channels, as shown in Figure 1a, all of which are read out synchronously. Each channel has several regions available to collect calibration, or "collateral" data (Figure 1b). There are two sets of columns of virtual pixels: (1) 12 columns of bias-only pixels resulting from 12 leading pixels in the serial register ("leading black"), and (2) a 20 column serial over-scan region ("trailing black"). There are also two sets of rows of collateral pixels: (1) the first 20 rows, which are covered by an aluminum mask ("masked smear"), and (2) a 26 row parallel over-scan region ("virtual smear"). During science data collection, a coadded sum of specified columns of the trailing black and rows of both the masked and virtual smear are stored at each cadence for each channel. To enable correction for some of the artifacts a specific set of artifact removal pixels (ARPs), as illustrated in Figure 1c, are collected along with the science data.

\*kolodz@nasa.gov; phone 1 256 961-7775; fax 1 256 961-7522; nasa.gov

Science data are available at either short cadence ( $\sim 1$  minute) for 512 targets, or long cadence ( $\sim 30$  minutes) for 170,000 targets. All science data are collected with an integration time of 6.02 s with pixels read out at a 3 MHz clock rate. In science collection mode, the full single integration CCD frames are coadded together. Then at the end of the short and long cadence period pre-specified pixels for each target are selected from the coadd, processed, and stored on board. Due to data storage and transmission limitations, only about 6% of the 96 million pixels are stored for eventual transmission to the ground.

*Kepler's* shutterless operation precludes standard dark frames. Instead, reverse-clocking of the CCDs permits us to measure the bias level throughout the image in the absence of sky signal. Also, a full frame image (FFI) mode permits collection of all the pixels in the focal plane. FFIs are used to examine detector properties, verify pointing, and verify the target aperture definitions. Reverse-clocked data and FFIs are taken periodically throughout the mission.<sup>7</sup>

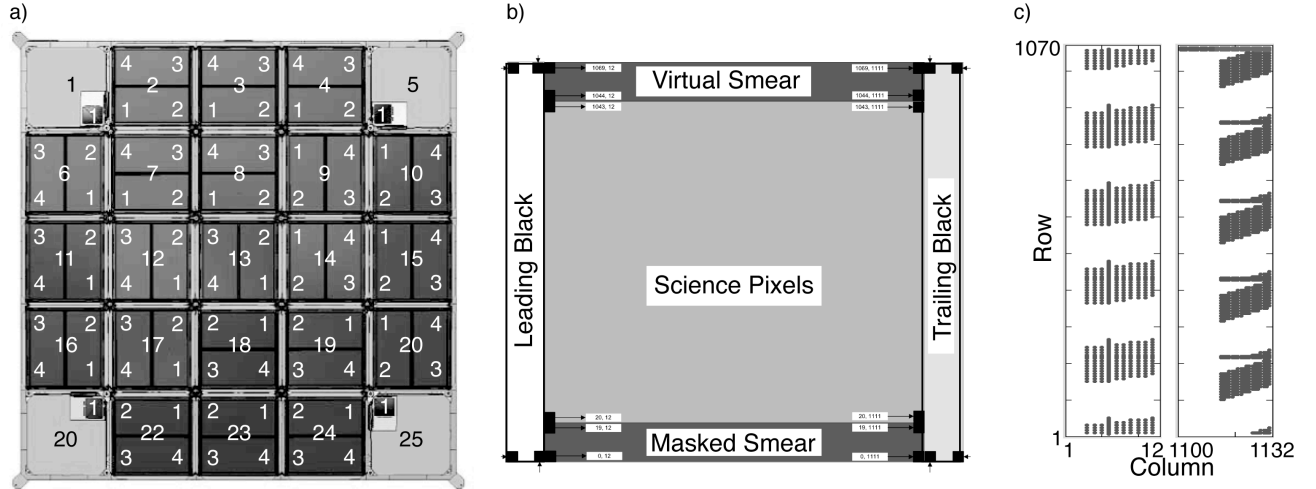


Figure 1. a) Focal plane with designations for modules 1-25, and outputs 1-4 for each module. Module outputs 1.1, 5.1, 20.1 and 25.1 are fine guidance sensors (FGS). b) Identification of collateral regions associated with each output channel. c) Plots indicating the locations of artifact removal pixels (ARPs) in the leading and trailing black regions. These pixels are collected during each long cadence.

## 2. OBJECTIVES

To illuminate the motivation for this activity we begin with a brief description of the artifacts of concern. We identify the *Kepler* noise floor and relate this to the pattern noise levels or bias variability induced by the artifacts, defining an acceptable level for the bias variations. Our flagging and removal objectives are based on these levels.

### 2.1 Description of artifacts

Ground testing uncovered several instrumental artifacts, each of which was investigated to understand the cause, impact, and cost to fix or mitigate. These artifacts were extensively characterized on the ground and then again during commissioning. These investigations determined that several did not require mitigation. The existing data processing pipeline<sup>6</sup>, or the algorithms described herein handle those with the largest impact. Those relevant to this paper are briefly described below and illustrated in Figure 2.

**FGS Clocking Cross Talk.** Cross talk from the FGS clocks to the science CCD video signals injects a complex pattern into the bias image of every science channel with an amplitude up to 20 DN read<sup>-1.3</sup>. Because the FGS and science CCDs share the same master clock, the pattern is spatially fixed; however, the amplitude of the cross talk is dependent on the temperature of the Local Detector Electronics (LDE). The cross talk has three distinct components based on the state of the FGS CCDs as the science pixel is read out (see Figure 2a): FGS CCD frame transfer, parallel transfer, and serial transfer. Approximately 20% of targets have at least one of the parallel or frame-transfer cross talk pixels in their aperture. Without mitigation, the cross talk introduces a small time-varying bias into a target's flux time series as the LDE temperature changes.

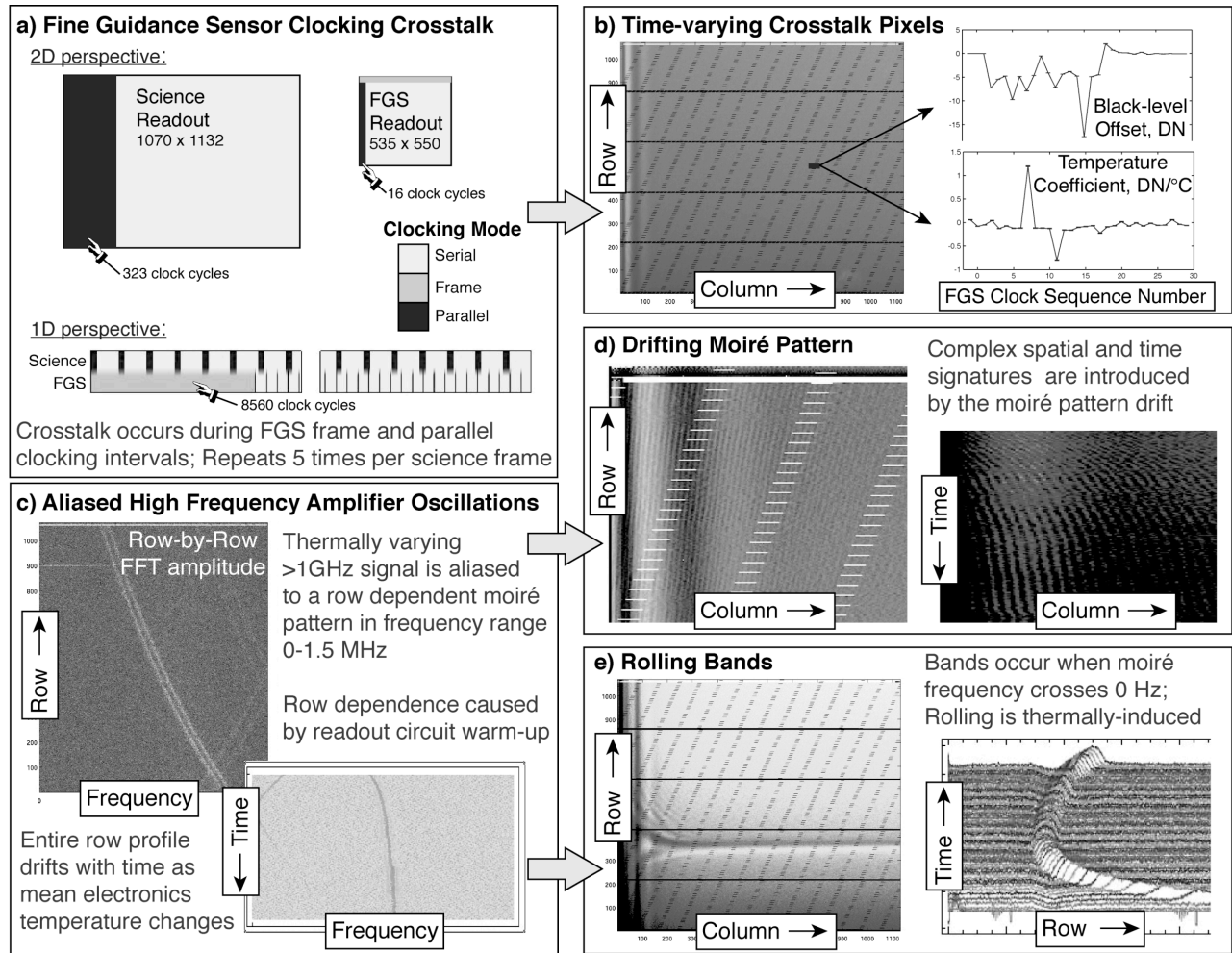


Figure 2. Illustrative description of pattern noise sources (left) and effects on *Kepler* images (right). a) The FGS pixels are read out synchronously with the science signals but the difference in size of the two sensors combined with the changes in characteristics of the clocking signals during the parallel and frame transfer intervals produces b) a complex pattern of thermally varying pixels at specific locations across the image. c) An aliased high frequency signal is highly sensitive to LDE component temperature changes due to normal operational warm-up. These produce a smoothly varying frequency change with row. Note that a 0.1% change in the source frequency can produce an aliased frequency change from DC to Nyquist (1.5 MHz) in the readout. One result is d) a moiré pattern with row-dependent frequency (left). The trace of a single row with time shows drift in the frequency (right). The other result is e) bands which occur when the aliased frequency crosses 0 Hz, extend over all columns (left). The time-evolution of a single column shows rolling band in response to temperature changes (right). Plots are meant to show typical qualitative behavior for affected channels, so scales and data sources are suppressed. For reference b) shows flight data from output 20.2, c) shows ground test data from output 14.4, d) shows ground test data from output 9.3 and e) shows ground test data from 13.4.

**High-frequency Oscillations.** A temperature-sensitive amplifier oscillation at >1 GHz was detected in some CCD video channels during the artifact investigation. Investigation suggests that this signal may originate is from the AD8021 operational amplifiers used extensively in the video signal chain, which may show subtle layout-dependent instability when used at low gains. The oscillation's frequency range, rate of change, and pattern among the channels matched closely those characteristics in the dark images, strongly suggesting that the artifact is a moiré pattern (MPD) generated by sampling the high-frequency oscillation at the 3MHz serial pixel clocking rate. Since the characteristic source frequency drifts with time and temperature of the electronic components by as much as 500 kHz/°C, the signal from a given pixel in a series of dark images has a time varying signature. This signature may be highly correlated with neighboring pixels and yet poorly correlated with slightly more distant pixels. When the oscillation frequency is a

harmonic of the serial clocking frequency, a DC shift occurs producing a horizontal band offset from the mean bias-level in the image. As the frequency drifts with temperature, the point on the image where this DC shift occurs moves up or down from sample-to-sample, producing a rolling band (RBA).

Forty-six of the 84 readout channels have never exhibited moiré pattern behavior and an additional 9 channels have thus far not exhibited this behavior at a detectable level in flight. Ten channels have shown rolling bands. While the moiré amplitude per pixel in the remaining channels is significant, its affect on our ability to detect small planets depends on frequency, sum within a target aperture, and variations over time-scales of interest to transit detection. Based on the first 33.5 days of data from science operations, the instrument is meeting the 6-hour precision requirement across the focal plane for the quietest 30% of stars.<sup>9</sup> The two worst moiré channels, 9.2 and 17.2, exhibit a ~20% increase in 6-hour noise over the focal plane average at 12th magnitude, based on the standard deviation of 6-hour binned flux time series. Such an increase is small compared with the factor of 1.5 spread in the distribution of dwarf star precision at 12th magnitude.<sup>9</sup>

For completeness, we briefly mention several other instrumental features in the *Kepler* data which are either accounted for in the current calibration scheme<sup>6</sup> or are not currently observed to be variable enough to adversely affect *Kepler* science. The list includes: 1) *Time-varying low-spatial-frequency characteristics*, whereby the row-profile of the 2D-black image tends to evolve slowly with time. 2) *LDE Undershoot*, whereby star-like images induce a signal-dependent trailing undershoot in the video output, and 3) *Start-of-line Ringing (SOLR)*, whereby a transient signal is initiated at the onset of serial clocking of each row. The variations at low spatial frequency are corrected in the current calibration using a cadence-by-cadence fit to the trailing black collateral data. We use a similar fit in the algorithm described in section 3 which also accounts for FGS crosstalk and LDE undershoot. This algorithm is designed to replace the current calibration method in a future pipeline release. An undershoot correction is also included in the current calibration pipeline with a model including up to 20 pixels when they are available in the target table. We have observed, and continue to expect the start-of-line ringing to be sufficiently stable over time to avoid photometric precision degradation.

Finally, *scene dependent artifacts* arise in two possible ways: 1) as a consequence of the sensitivity of the oscillating LDE component to temperature, the thermal transient introduced during readout by the signal from a bright star causes additional localized changes in the detected moiré pattern amplitude and frequency, and 2) time variability of pixels detecting bright stars, especially at the ends of saturated column segments, the LDE undershoot from beyond 20 pixels or from undetected pixels causes time variability in the black level. Section 5.3 discusses how we will address these in future development if we find that they warrant high enough priority relative to other factors influencing photometric precision.

## 2.2 *Kepler's* noise floor

The *Kepler* design placed the noise floor for detecting stellar variation at a level that enables detection of earth sized planets in habitable orbits around 12th magnitude stars. A remotely observed earth would produce an 84 ppm decrement during a 12 hr solar transit, and to ensure high detection efficiency for similar pairs of objects, a  $1\sigma$  noise level of 20 ppm for 6.5 hour periods was allocated. Artifact-induced changes in bias level significantly exceeding 20 ppm on this time scale or longer could be confused with the signals resulting from stellar variability. Therefore, the objective of this work is to correct bias variations to below the noise floor where possible, and to detect and flag them where correction is not possible. Table 1 summarizes implications for the acquired signal levels in electrons and DN.

Table 1. Quantities used to define the level at which artifacts begin to affect *Kepler* science.

Stellar magnitude (G-type star)	11	12	13
Detected in-aperture electrons in 6.5 hrs, aperture sizes are typical for the magnitude	1.12e+10 in 25 pixel aperture	4.47e+09 in 16 pixel aperture	1.79e+09 in 9 pixel aperture
Earth-equivalent transit decrement, 84 ppm ( $e^-$ )	939000	376000	150000
Allocated noise including shot noise, 20 ppm for 12th mag. ( $e^-$ )	168000	89400	51900
Bias change matching allocated noise level in aperture in 6.5 hr ( $e^- \text{ read}^{-1} \text{ pixel}^{-1}$ )	1.9	1.6	1.7
Bias in DN, nominal 100 $e^-$ /DN scale factor (DN $\text{read}^{-1} \text{ pixel}^{-1}$ )	0.019	0.016	0.017

The characteristics of pattern noise that justify the additional attention are that it may be spatially correlated and temporally correlated, whereas ordinary white noise or other broadband types of noise are likely to be less so. For an

aperture with 16 pixels, observed for a period of 6.5 hours (13 long cadences) noise will grow in the sum at a rate  $\sqrt{16 \times 13} \approx 14$  times more slowly than a spatially and temporally correlated bias change. On the other hand, pattern characteristics vary widely, so only a small fraction of the integrated observing time  $\times$  field-of-view product is likely to be subjected to a reduction in sensitivity or increased likelihood of false detection. In this paper we use the term "source coverage" to identify the product of time  $\times$  field-of-view.

### 2.3 Acceptable bias variations

It is clear that space and time correlated signals below 0.02 DN/read over time intervals shorter than 6.5 hours would be difficult to distinguish from noise and therefore have little scientific impact. Typically, the acceptable level of bias variation would be constrained by detectability to several sigma above the noise level. In this case however, we are able to leverage from the similar behavior of pixels acquired at a specific FGS clocking interval or pixels in an extended region of a given readout channel to measure bias variations at several times below the noise level. Since it is no more difficult to observe effects at the noise level than several times above, we simply round the 0.016 DN read<sup>-1</sup> pixel<sup>-1</sup> to 0.02 DN read<sup>-1</sup> pixel<sup>-1</sup> which is equivalent to 25 ppm of the 6.5 hr. signal from a 12th magnitude star. Thresholds for correction and flagging of artifact-induced bias variations are based on limiting changes to 0.02 DN read<sup>-1</sup> pixel<sup>-1</sup>.

### 2.4 Removal and flagging objectives

FGS crosstalk is clearly detectable in data from collateral regions collected during every long cadence. The multiple examples of pixels collected during each FGS parallel and frame clocking interval typically indicate a repeating pattern which shows little or no change over one science readout. It is therefore possible to measure these bias changes on a cadence-by-cadence basis and have high confidence that the science pixels are subject to the same effects. *The objective for the FGS algorithms is therefore to remove the FGS crosstalk signal to the level of 0.02 DN read<sup>-1</sup> pixel<sup>-1</sup>.*

Collateral data from smear regions also provides a way to measure the amplitude and frequency of moiré pattern on a cadence-by-cadence basis at the beginning and end of each readout interval. Relating this to the row-by-row amplitude and frequency of moiré pattern in the difference between FFIs provides an indirect means of estimating the characteristics of moiré pattern at any readout location and time. Changes in trailing black collateral provide a similar means to detect and characterize rolling bands. For these however, confidence in the exact estimate of the artifact-induced bias does not match that of FGS corrections. Such efforts are complicated significantly by complex phase variations and scene-induced changes in readout component temperatures, which lead to short-term changes in the oscillating frequencies and local shifts in the moiré patterns. *For these reasons, our objectives with respect to MPD and RBA are to flag regions of the focal plane and times when collateral data indicates that the amplitude of an effect is greater than 0.02 DN read<sup>-1</sup> pixel<sup>-1</sup>.* The impact of a specific moiré pattern amplitude on the peak-to-peak variation in an aperture is reduced by a factor  $|2 \sin \pi f n / \pi f n|$ , where  $n$  is aperture width in pixels, and  $f$  is the moiré spatial frequency.

This makes it difficult to generalize the severity of the moiré pattern in a flagged region independent of the target apertures. Our approach is to flag data based on amplitude and provide localized severity information, such as frequency, which permits more precise evaluation of the impact on a given target.

## 3. METHODS

The process developed to mitigate the effects of the pattern noise sources on *Kepler* science includes the elements and data products shown in figure 3. The overall architecture takes advantage of the results of cadence-by-cadence spatial fitting, removing FGS crosstalk and low-spatial frequency variations from fit residuals, which would otherwise complicate detection of the rolling bands and moiré pattern.

### 3.1 Spatial fitting

The spatial fitting algorithm is designed to extract information about the time-varying parts of the *Kepler* data stream using pixel and collateral data from each cadence or FFI. We derive the information in the form of fit coefficients and uncertainties based on a model of the observed behavior of each pixel. Let  $X_{RC}$  represent the raw black level value in DN at row  $R$  and column  $C$ , and let  $Z$  represent the zero offset introduced to prevent negative values. Then we assume,

$$X_{RC} - Z = X_0 + f_{RC} + g_C + W_{RC}^F + W_{RC}^P + U_{RC}$$

where,

$X_0$  is a constant

$f_{RC}$  is the row dependent part, (C only distinguishes leading and trailing black)

$g_C$  is the column dependent part

$W_{RC}^F$  is the FGS frame crosstalk dependent part

$W_{RC}^P$  is the FGS parallel crosstalk dependent part

$U_{RC}$  is the undershoot dependent part

Each of these components is defined in the Appendix in terms of linear coefficients. The fit values of all the coefficients, coefficient errors, residuals and associated statistics are the data products of this algorithm.

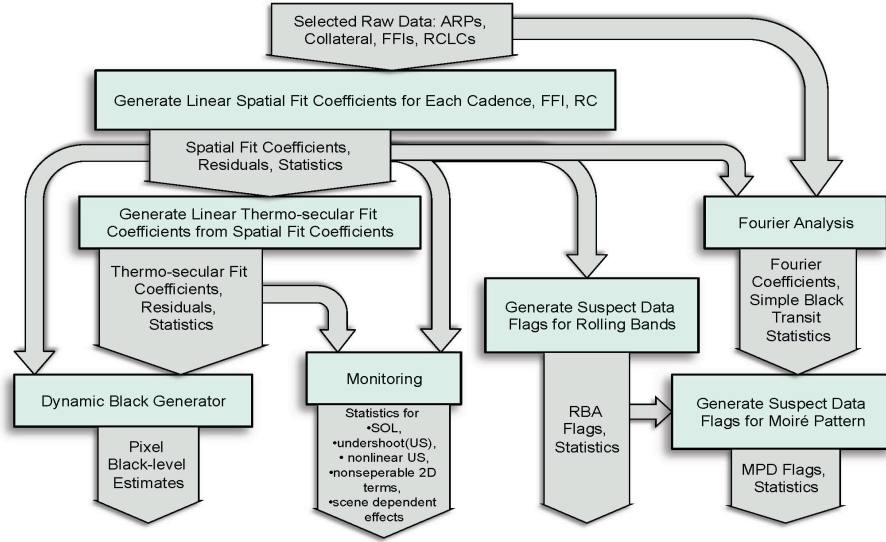


Figure 3. Architecture diagram showing prototype elements as boxes and data product flow as arrows.

The process includes initialization and fitting segments. Initialization of FGS clock states as a function of pixel location, selection of scene dependent exclusion zones, initialization of model components are parts of the initialization segment. Fitting the average of specified cadences and fitting each of a series of single cadences comprise the fitting segment.

The FGS frame and parallel crosstalk components occur in repeating patterns defined by the parallel and frame clocking intervals of the of the FGS. The term "row" applies to one of 1070 rows of a science channel image, each 1132 pixels long. The actual number of 3MHz clock cycles in a row is 1455, so 323 clock cycles don't result in pixels (science parallel clocking). There are 5- 311370 clock-cycle FGS frames for every science frame, so an overall pattern repeats every 214 rows. In terms of the continuous clock stream, the frame crosstalk signal repeats every 16 pixels for 8450 clock cycles. It then flat-lines to a constant offset for 109 clock cycles followed by a short 3-clock cycle waveform. The parallel crosstalk signal repeats every 566 clock cycles with the first 16 cycles having the strongest crosstalk signal followed by a another 12-24 cycles which exhibit low-level "ringing". The remainder of the 566 may be treated as constant except the last one which shows variability in some channels (cycles 17-566 are the serial clocking intervals of the FGS frame readout). The parallel pattern begins at FGS clock cycle 8563 with parallel cycle number 3 of 566 and continues modulo 566 through the entire FGS clock cycle range to 311370.

To serve the fitting algorithm we assign the FGS-frame clock sequence numbers 1-16 to the repeating frame pattern, the value 17 to the 109 following pixels and 18-20 to the interval-ending feature. The remainder of the 311370 are assigned the value 0. Similarly, we assign the FGS-parallel clock sequence numbers 1-566 to the repeating parallel pattern starting with value 3 assigned to the 8563rd cycle, and all preceding cycles are assigned the value 0.

After stacking the 5 FGS frames and chopping off the excess 323 unobserved columns, this provides a pair of mappings from (row, column) to FGS-frame clock sequence number, and (row, column) to FGS-parallel clock sequence number. We fit a spatial model in which each pixel's black level is a separable function of row, column, FGS-frame clock sequence number (frame-CSN), and FGS-parallel clock sequence number (parallel-CSN). The algorithm allows fitting

any combinations of frame-CSN and parallel-CSN, but we have limited the parallel-CSN modeling range to pixels 1-29 and 565-66. The range 30-564 does not appear to contain time varying features. The frame-CSN pixels 18-20 are only measured in reversed-clocked cadence data, and are monitored but assumed constant.

Regions are excluded from the fits because they are potentially affected by moiré pattern drift changes or undershoot produced by signals from bright stars near the trailing black. These appear to be caused by temperature changes induced by the signals from bright stars on the readout circuit. For brightest stars, the signal may bleed over many rows and therefore the scene dependent region extends over many rows. The algorithm used for selection of scene dependent exclusion zones uses FFI images, and depend on choices of: a) a signal threshold in DN read<sup>-1</sup>, above which a pixel may cause a scene-dependent artifact, b) a column threshold, above which a signal-threshold-crossing pixel may cause a scene-dependent artifact in the trailing black, and c) the row pad count to extend each contiguous region to account for potential variations in star signal levels. Essentially, if a row of a robust averaged (cosmic ray clean) FFI contains a pixel above the signal threshold at a column beyond the column threshold, then any row within the pad count of that row will be excluded from the trailing black fits.

Only specific regions of an image are suitable for determining spatial coefficients. To systematize the selection of these regions the algorithm accepts specifications of a series of rectangular regions of interest (ROI) which can be thought of as building blocks for the complete modeled image region, containing pixel or collateral point attributes needed to assemble the modeled response vectors and design matrices. These are summarized in Table 2.

Table 2. Regions of interest containing data used for black level spatial fitting. The 'type' column simply identifies the source of response vector data, "ARP" means artifact removal pixels from target LC data, "Collateral" means summed collateral data, "FFI" means FFI pixel data, "RCLC" means target pixel data from reverse-clocked long cadences. The target table which specifies RCLC pixels was designed to monitor the FGS crosstalk, RBA and MPD artifacts.

Region of Interest (ROI)	min row	max row	min col.	max col.	type	main use
Leading ARP	7	1059	3	12	ARP	<ul style="list-style-type: none"> <li>• row dependence</li> <li>• FGS crosstalk</li> <li>• undershoot</li> </ul>
Trailing ARP	7	1051	1115	1132	ARP	
Trailing ARP Undershoot	1052	1063	1113	1132	ARP	
Trailing Black Collateral	7	1059	1119	1132	Collateral	
Trailing FFI	7	1063	1113	1132	FFI	<ul style="list-style-type: none"> <li>• column dependence</li> <li>• SOLR</li> <li>• FGS crosstalk</li> </ul>
Masked Smear Collateral	7	18	13	1112	Collateral	
Virtual Smear Collateral	1047	1058	13	1112	Collateral	
Reverse-Clocked Long Cadence	7	1058	3	1130	RCLC	

The complete model is assembled from a set of components. To systematize the construction of these components we identify represent distinct namable terms in the model which can be thought of as building blocks for the complete modeled design matrix, and associated indices for exclusion of potentially scene-dependent rows as a function of channel and all-zero columns in the design matrix. These are essentially the components of  $X_{RC}$  described at the beginning of this section. There are also "delta" components, which measure the difference between leading and trailing black or masked and virtual smear as detailed in the appendix. The design matrix is constructed with one row of information for each element of acquired data within each ROI model.

Prior to performing the cadence-by-cadence fitting for a given channel, we fit the pixel-by-pixel average of all specified cadences to obtain the exponential time constant parameter for the row-dependent exponential term using a nonlinear model and then perform a linear fit on the mean of pixel values over selected long cadences to produce a set of mean coefficients. The resulting exponential time constant is used for all single-cadence fits for a given channel.

Finally, we perform cadence-by-cadence fits to obtain linear coefficients for individual cadences. The algorithm consists of the following steps:

- For each channel and for each long cadence:
  - Select response vector data from input raw long cadences using the ROI indices described above.
  - Determine the cadence-specific undershoot component of the design matrix, which we apply to only a small subset of the modeled pixels as described in the appendix.
  - Concatenate the constant part of design matrix with the undershoot component.
  - Perform a linear fit of scene-dependent-free data to obtain linear coefficients, and store results in output structures.

- For each channel and for each FFI:
  - Select response vector data from input raw FFI using the ROI indices described above.
  - Determine the FFI-specific undershoot component of the design matrix, includes all modeled pixels.
  - Concatenate the constant part of design matrix with the undershoot component.
  - Perform linear fit of all data to obtain linear coefficients, and store results in output structures.

### 3.2 Thermo-temporal fitting

The spatial fit for each cadence produces a time-series of spatial coefficients. These coefficients exhibit a variety of behaviors, which are often, but not always, simple functions of time or temperature. The purpose of the thermo-temporal fitting algorithm is to fit the time series of each spatial coefficient,  $C$ , to the equation:

$$C(t, T) = K_0 + K_t t + K_T T$$

where,

$K_0$  is a constant

$K_t$  is the linear trend in  $C$

$K_T$  is the temperature coefficient of  $C$

$t$  is time

$T$  is temperature.

In addition the algorithm provides the necessary statistics to evaluate whether the model fit is consistent with any of the possibilities  $K_t=0$ ,  $K_T=0$  or both  $K_t$  and  $K_T=0$ . This evaluation is based on the  $\chi^2$  of the four fit combination with zero or nonzero  $K_t$  or  $K_T$ .

### 3.3 2D-black correction

Not all the information from the spatial and thermo-temporal fits is typically required to adequately correct the data for the observed artifact-induced variations. Some of the fit parameters are intended only for monitoring instrument performance. The 2D black correction only includes the following terms.

$$X_{RC} - Z = X_0 + f_R + g_C + W_{RC}^F + W_{RC}^P$$

where,

$X_0$  is a constant

$f_R$  is the row dependent part

$g_C$  is the column dependent part

$W_{RC}^F$  is the FGS frame crosstalk dependent part

$W_{RC}^P$  is the FGS parallel crosstalk dependent part

The undershoot and the "delta" components are treated as static, and monitored. In the event that time variability in these terms becomes significant enough in the future to impact science results, these terms could be added to the correction algorithm. The equation for  $X_{RC}$  is separable in terms of the row, column and crosstalk dependent parts, which enables the 2D corrections to be calculated from only the four vector terms, significantly shorting the processing time compared with an inseparable function.

The algorithm is implemented as a function with arguments 1) a set of pixel rows, 2) a set of pixel columns, 3) a set of long cadence numbers, which specify time, and 4) a parameter that determines whether the row and column lists should be interpreted as defining a discrete set of points or a rectangular region. The model is assembled from the spatial and thermo-temporal fit coefficients based on a decision tree. If the  $\chi^2$  of the thermo-temporal fit indicates that the time series of the given spatial coefficient is not excluded from being a specific model at the 95% confidence level, then the spatial coefficient is determined from that model for all specified LC. In the event that multiple models meet the criterion, the preference order for the models, from highest to lowest is: constant, constant+linear-thermal, constant+linear-trend-with-time, constant+linear thermal +linear-trend-with-time. If none of the cases meet this criterion, a comparison between the standard deviation of the differences between consecutive coefficients in the time series and the spatial coefficient fit errors is made. If the standard deviation of the differences is more than 1.5 times larger than the standard errors in the individual coefficients, as determined in the least-squares spatial fits, then the coefficients are applied discretely, otherwise they are smoothed using a quadratic thermo-temporal fit over an adaptive time interval.



### 3.4 Flagging

The RBA and MPD share the same source signal, the high frequency amplifier oscillation described above, however RBA are somewhat simpler to identify. The effects of RBA on pixel time series are generally larger and may be a greater risk to complicate the search for planet signatures. For these reasons, we have developed separate algorithms for detection and flagging of each.

Overall datasets are initially divided into subsets defined by a range of rows, a range of columns, and a range of long cadences. Both algorithms share the same general flow involving the following steps:

*Measurement:* A measurement algorithm operates on each data subset to determine parameters indicating the presence of unusual bias level behavior. The accumulation of these parameters over the entire data set provides input to detection algorithms.

*Detection:* Detection algorithms compare the measured parameters with acceptability thresholds to detect unacceptable artifact behavior. The result is a map determining which parts of the full data set exceed the threshold criteria.

*Postprocessing:* Postprocessing algorithms accomplish two main tasks, a filter to clean spuriously flagged regions and a padding algorithm to add a buffer zone around areas with a high density of above-threshold flags. The former prevents unnecessary flagging of data which is not at significant risk to interfere with planet search algorithms. The later accounts for the fact that we do not directly measure the science pixels but rather collateral data which is indicative of the science pixels, and thus there is some level of uncertainty in boundaries of affected locations. A final step in the postprocessing identifies rectangular boundaries in the padded flagged regions to be listed as suspect data flag objects to be carried forward as input to pipeline algorithms or as data products supplied to users.

*Severity evaluation:* A final algorithm assembles a set of statistics which characterize the severity of the flagged artifacts in the identified regions. The algorithms operate on the data corresponding to the boundaries derived for the suspect data flag objects to produce a set of parameters describing the pixels and cadences being flagged by that object. The flags are paired with the severity parameters to form a list which represents the primary output of the flagging algorithms. The distinctions between the rolling band and moiré pattern algorithms are summarized in the Table 3 with further detail in the ensuing paragraphs.

Table 3. Flagging algorithm implementation comparison between rolling band flagging and moiré pattern drift flagging. Numerical values specified are nominal values for algorithm parameters.

Item	RBA	MPD
Measurement	<ul style="list-style-type: none"><li>time-varying displacement in trailing black spatial fit residuals, not due to scene dependent artifacts</li><li>black transit search in trailing black residuals and smear differences</li></ul>	<ul style="list-style-type: none"><li>Amplitude of moiré pattern in 19 110-column subregions</li><li>frequency of moiré pattern</li><li>black transit search in trailing black residuals and smear differences</li></ul>
Detection	<ul style="list-style-type: none"><li>displacement sigma vs. time &gt; 0.02 DN/pixel/read</li><li>-or-</li><li>both fixed residual displacement and black transit &gt; 0.02 DN/pixel/read</li></ul>	<ul style="list-style-type: none"><li>moiré pattern amplitude &gt; 0.02 DN/pixel/read in subregions</li><li>-or-</li><li>black transit &gt; 0.02 DN/pixel/read</li></ul>
Postprocessing	<ul style="list-style-type: none"><li>filter out flag densities less than 5% in 10 row by 39 LC region</li><li>pad around surviving flags <math>\pm 19</math> LC and <math>\pm 10</math> rows</li><li>package flagged regions into rectangular suspect data flags (SDF).</li></ul>	<ul style="list-style-type: none"><li>filter out flag densities less than 5% in 10 row by 39 LC region</li><li>pad around surviving flags <math>\pm 18</math> LC and <math>\pm 10</math> rows</li><li>package flagged regions into rectangular suspect data flags (SDF).</li></ul>
Severity Evaluation	<ul style="list-style-type: none"><li>calculate statistics of measured parameters within each SDF region.</li><li>97.7 percentile &amp; median displacement and fraction of exposure exceeding displacement</li></ul>	<ul style="list-style-type: none"><li>calculate statistics of measured parameters within each SDF region.</li><li>number of traceable moiré spectral peaks detected</li></ul>

	<p>threshold</p> <ul style="list-style-type: none"> <li>• 97.7 percentile &amp; median noise and fraction of exposure exceeding noise threshold</li> <li>• 97.7 percentile &amp; median bias variation and fraction of exposure exceeding bias variation threshold</li> <li>• number of &gt;3 &amp; &gt;4 sigma transit-like features</li> <li>• total fraction of exposure exceeding combined thresholds</li> </ul>	<ul style="list-style-type: none"> <li>• 97.7 percentile &amp; median peak-to-peak moiré spatial amplitude and fraction of exposure exceeding amplitude threshold for each trace</li> <li>• median spatial frequency at peak amplitude for each trace</li> <li>• maximum temporal amplitude and corresponding frequency</li> <li>• 97.7 percentile &amp; median noise and fraction of exposure exceeding noise threshold</li> <li>• 97.7 percentile &amp; median bias variation and fraction of exposure exceeding bias variation threshold</li> <li>• number of &gt;3 &amp; &gt;4 sigma transit-like features</li> <li>• total fraction of exposure exceeding combined thresholds</li> </ul>
--	--	---

*RBA.* The signature of a rolling band is a time-varying displacement in trailing black spatial fit residuals, not due to scene dependent artifacts. The algorithm searches for these on a row-by-row basis for each channel. It also convolves a square wave transit kernel with these time series, as well as the column-by-column difference between masked and virtual smear time series for places where the black level variations exhibit transit-like signature. We call these "black transits." Clearly, when the collateral data exhibits transit-like time signatures, the data should be flagged, even if the typical variation is small. The convolution is equivalent to a least-squares fit to a transit signature at centered at each long cadence. The fit transit depth is treated as a bias, and the uncertainty in the transit depth determines the noise level.

*MPD.* The MPD flags require additional processing steps which include fast Fourier transforms of both the difference between a pair of FFIs and the difference between masked and virtual smear collateral data for each long cadence. The moiré pattern produces a detectable trace of high amplitudes with a continuous frequency signature as shown in figure 8. The FFI gives a spatial distribution of moiré amplitudes while the smear differences give variation in time. By matching the smear frequencies in the FFIs to the nearest long cadence and then scaling the spatial distribution on a cadence-by-cadence basis we can estimate the moiré pattern amplitude in any row at any time. FFTs are also applied to 19 110-column-wide subregions at 55 column steps to measure the column dependence of the amplitude within a narrow band of the frequency determined from the full-width image. Threshold-crossing of amplitudes from the windowed regions identifies flagged region boundaries.

Often, as in figure 8, multiple frequency traces appear in the FFI difference row-by-row FFT or masked-virtual smear difference cadence-by-cadence FFT. These must be resolved to correctly determine the frequency of the maximum moiré pattern amplitude. A brief outline of the algorithm that performs this task includes the following steps.

- Determine the 4 largest FFT amplitude peaks separated by some minimum frequency difference in each row of the FFIs and in each long cadence of the smear difference time series.
- Determine the slope (Hz/row or Hz/long cadence) and intercept (Hz) of a line through each pair of peaks in the set which are separated by less than a specified maximum number of rows or long cadences.
- Construct a 2D histogram of pair occurrences binned by slope and intercept.
- Identify contiguous high-occupancy regions, or islands in the 2D histogram and establish a threshold level which defines the boundaries of the islands. Each island corresponds to a unique frequency trace.
- Reconstruct each frequency trace by assembling a set of line segments of specified length using the slopes and intercepts only from pairs of peaks which occupy the island. Each segment is centered at the mean frequency and mean row or long cadence of the pair.
- Obtain frequency versus row or long cadence by averaging the segment frequency values where they intersect a given row or long cadence number.

The final step determines the amplitudes associated with each detected frequency trace and column subwindow as a function of row or long cadence. The FFIs and smear differences give amplitude and frequency as a function of row,  $r$ , column subwindow,  $c$ , and time,  $t$  defined as:  $A_i^{FFI}(r, c)$ ,  $f_i^{FFI}(r, c)$ ,  $A_i^S(r_s, c, t)$ ,  $f_i^S(r_s, c, t)$ , respectively. Here the

index,  $i$ , references a specific matched frequency trace, and  $r_S$  is the associated smear row. The final amplitude function,  $A_i$ , and frequency function,  $f_i$ , for the  $i$ th frequency trace then follow from:

$$A_i(r, c, t) = A_i^S(r_S, c, t) \frac{A_i^{FFI}(r, c)}{A_i^{FFI}(r_S, c)}$$

$$f_i(r, c, t) = f_i^S(r_S, c, t) + f_i^{FFI}(r, c) - f_i^{FFI}(r_S, c)$$

The maximum amplitude among the various traces is the flag discriminator. The characteristics of all the detected frequency traces are summarized in the severity parameters for the flagged object.

## 4. RESULTS

We applied these algorithms to a cross-section of *Kepler* flight data. The following paragraphs describe the results of these prototype runs. Table 4 identifies the flight data used to exercise each algorithm. The channels 2.1 and 12.1 were selected as "good channels", 6.2 and 20.2 exhibit worst case FGS crosstalk, and 9.2 and 17.2 exhibit worst-case MPD.

Table 4. Module Outputs and time intervals used to exercise the various algorithms. Q0 lasted 10 days during commissioning from May 1-11, 2009, Q1 lasted 34 days from May 13-June 15, 2009.

Item	All Long Cadences Q0-Q1	All LC first 10 days of Q1	Decimated set: every 10th LC for Q0; every 20th LC for Q2
Fitting	2.1, 6.2, 9.2, 12.1, 17.2, 20.2	All	All
Correction	2.1, 6.2, 9.2, 12.1, 17.2, 20.2	None	None
RBA Flagging	2.1, 6.2, 9.2, 12.1, 17.2, 20.2	All	None
MPD Flagging	2.1, 6.2, 9.2, 12.1, 17.2, 20.2	All	None

### 4.1 Example fits

*Kepler* readout channels exhibit a wide variety of black-level morphologies. This section illustrates the model response fidelity using a few typical channels which span this range. Figure 4 shows the raw trailing black collateral data for two channels, 12.1 and 20.2 along with the fit curves. Some row regions were excluded from the fits because bright stars near the trailing black would have introduced unmodeled scene dependent artifacts that could bias the fit coefficients if not explicitly removed.

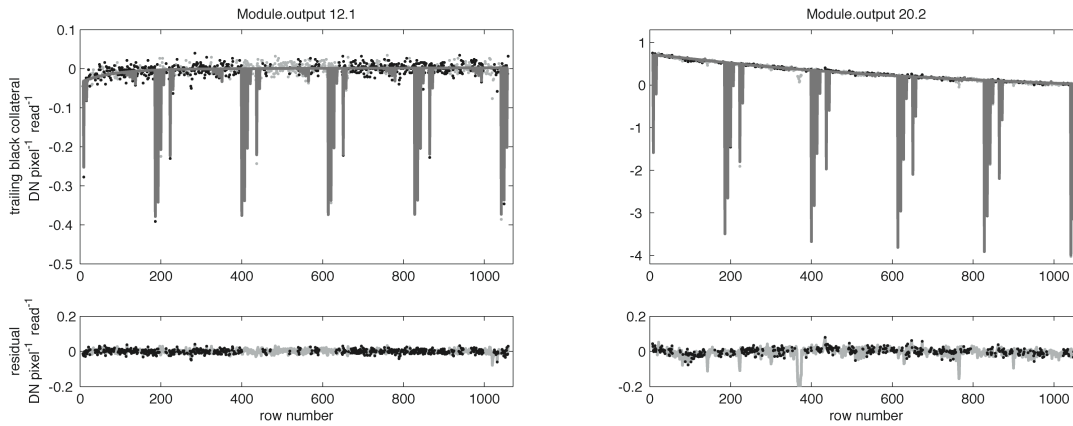


Figure 4. Example comparison of fit curves (dark gray) and data (black) comparing trailing black collateral for channels 12.1 and 20.2 for one representative Q1 long cadence. The scales include an arbitrary constant offset. The row dependence of the black level is modeled well by an exponential term with time constant which varies from channel to channel plus a logarithmic term which is channel independent. The light gray points are data and residuals from regions excluded from the fit because of increase likelihood of scene dependent bias due to stars with pixel values  $>5000$  DN  $\text{read}^{-1}$  within 400 columns of the trailing black in the excluded rows. The density of stars is higher in channel 20.2 in Q1, so the likelihood of stars very close to the trailing black is higher, leading to the evident higher number of obvious outliers in the excluded region in that channel.

The variation in the low frequency row-dependent terms of the model is evident from the two examples above. The series of spiked rows, repeating 5 times, are collateral rows containing FGS crosstalk sensitive pixels. The fit estimates offset values for each crosstalk-sensitive pixel from the combination of both the collateral row values representing the sum of 14 pixels and the individual ARPs. Undershoot coefficients represent filter coefficients for an additive term to each pixel based on a linear combination of the values of the previous 20 (not shown). Figure 5 shows typical time dependent behavior of low-frequency serial and FGS-frame coefficients. FGS-parallel coefficients are similar.

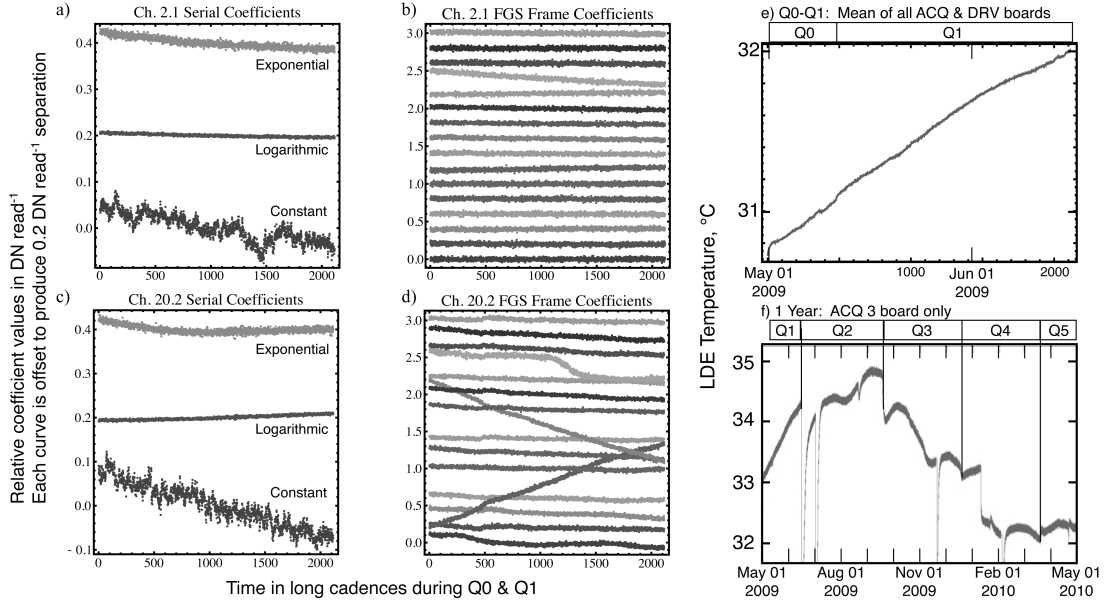


Figure 5. Examples of behavior of spatial fit coefficients vs. time for channels 2.1 (a,b) and 20.2 (c,d). a) and c) show the behavior of the low frequency spatial coefficients. The bottom traces show the constant term which is the only coefficient which always varies discretely (i.e. cannot be smoothed). The middle and top trace are the log and exponential coefficients respectively. b) and d) show the FGS frame crosstalk coefficients. The scales are DN/pixel/read vs. time as measured by long cadence number. An arbitrary offset has been added to each curve to provide visual separation. LDE temperature variation is shown in e) for Q0-Q1 only and f) for an entire year. The LDE temperatures in e) are derived from the mean of temperatures measured on all 5 LDE board pairs (ACQ & DRV), whereas f) shows only the ACQ 3 board temperature. Temperature and time were strongly correlated during Q0-Q1, but less so for the other quarters. Spacecraft rolls at quarter boundaries, except Q0-Q1, produce step changes, and the failure of module 3 in Jan-2010 produced a large step in ACQ 3, which is used for module 3 readout. Safe modes explain the remainder of the discontinuous features in f). Temperature variations are generally  $\sim 1^\circ\text{C}$  within a quarter, and  $\sim 2^\circ\text{C}$  over the year, excluding the step due to module failure.

#### 4.2 Effect of corrections on targets

Calibrated *Kepler* pixels are subject to 3 separate effects of FGS crosstalk stemming from a) the 1-D black correction, b) the smear correction and c) the effects of any FGS-cross-talking pixels in the target aperture. In the current crosstalk-unaware 1-D black correction, a robust polynomial fit is applied to the trailing black collateral data. Even with a robust fitting algorithm there is likely to be some time varying bias introduced by cross talking pixels with nonzero weights on the fit coefficients and thereby on the correction of each science pixel. 89% of collateral rows are unaffected by FGS-cross-talking pixels. The introduced bias in a given row is applied to uniformly to all science pixels in that row. The crosstalk-unaware smear correction is a cadence-by-cadence column-dependent correction based on measurements from the masked and virtual smear collateral regions. Each smear region is the sum of 12 rows. 57 % of collateral columns are unaffected by FGS-cross-talking pixels, and {23%, 3%, 12%, 5%} have {4, 3, 2, 1} modeled pixels in a given column. The introduced bias in a given column is applied to uniformly to all science pixels in that column. Both effects are reduced by the averaging and filtering afforded by the robust fits for the 1D black correction and the averaging over typically 24 rows for the smear correction.

The FGS-cross-talking pixels in the science pixel region represent the primary motivation for developing a dynamic 2D black correction. Target apertures currently have no time-varying correction for cross-talking pixels. 82% of targets

contain no FGS-cross-talking pixels. In this case the introduced bias in a given pixel applies to only science pixels in that target aperture. The overall impact on the affected 18% of targets is substantially reduced by averaging over the whole aperture, and by the fact that the crosstalk is somewhat smaller and changes sign for some pixels, thereby producing further dilution by averaging.

Figure 6 shows a histogram of the peak-to-peak dynamic 2D black corrections per target for Q1 where mean LDE temperature varied by  $1^{\circ}\text{C}$  over the full 34 day interval. The figure shows both a worst case channel, which is highly susceptible to crosstalk, and the cumulative impact on all channels. The gray histogram shows the separated effect of FGS-crosstalk in the collateral regions. The combined local pixels plus collateral (collateral only) affect 13% (5%) of targets by more than 0.02 DN/pixel/read in the worst case channel, while they affect only 4% ( $>1\%$ ) of all targets at this threshold level. The potential exists for a larger fraction targets to be affected by  $> 0.02$  DN/pixel/read for full quarters, but because the temperature variation is large during this period, these values are likely to be typical of a quarter. It is noteworthy that the worst case average temperature coefficient for a target is  $\sim 0.6$  DN/pixel/ $^{\circ}\text{C}$  implying that a  $0.03^{\circ}\text{C}$  change in LDE temperature with the appropriate time signature would be required to produce the signature of an earth-sized planet. Even then, the thermal excursion would need to repeat 3 times at regular intervals to produce a false positive. A more likely possibility is that, left uncorrected, the less predictably varying pixels, as shown in figure 5d, would reduce sensitivity to transits for limited periods in a small fraction of targets. Nevertheless, the increased risk of unanticipated consequences that these simple-to-remove artifacts could have on planet detection or other astrophysical applications for *Kepler* data warrants including this correction algorithm in the ongoing evaluation process for *Kepler* science pipeline enhancements.

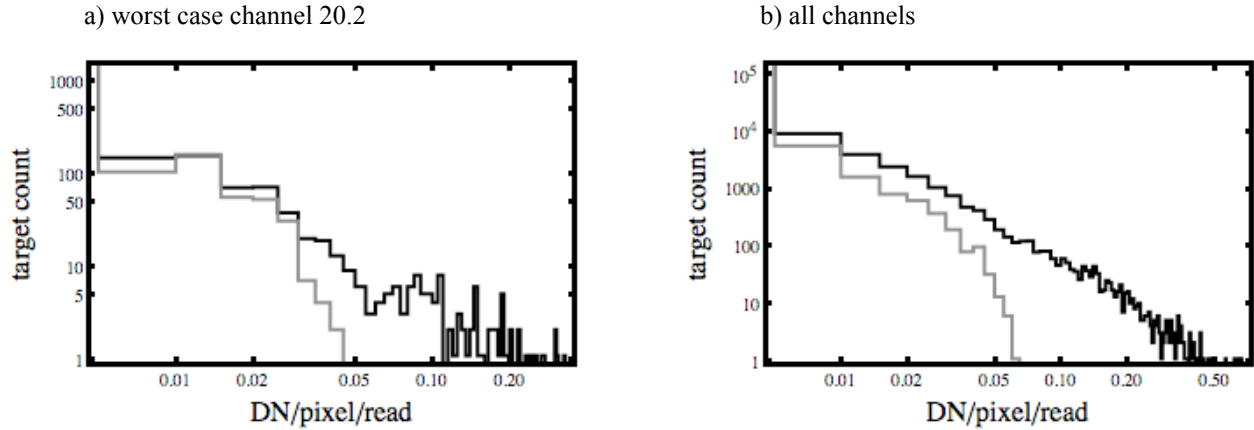


Figure 6. Histograms of peak-to-peak dynamic 2D black corrections per science target for a) worst case channel 20.2 and b) all channels during Q1. The contribution of collateral effects is shown in gray.

### 4.3 Flagging effectiveness

Two examples of rolling bands detected and flagged by the algorithm described above are illustrated in figure 7 for the result of combining quarters 0 and 1. The varying shaded areas represent the degree beyond threshold level that the data indicates a rolling band, the white regions are the padding around the offending regions and the black areas are unflagged. The figure shows the movement of the rolling bands as the flagged regions change with time. The pattern in the signal over time at the top of figure 7.b is from a scene dependent region caused by a variable star near the trailing black region. When a scene dependent region is flanked by a rolling band, the flagging algorithm automatically flags the scene dependent region as part of the rolling band. If no rolling band is present, the flag remains scene dependent. The rolling band flags apply to all columns in a flagged row so we can define the rolling-band-free source coverage as the fraction of black cells in these 2D maps. Here we define "artifact-free source coverage" as the fraction of available field-of-view solid angle  $\times$  exposure time that is unaffected by a given artifact. In the test cases below the rolling-band-free source coverage is 60% and 78% for module outputs 9.2 and 17.2 respectively. The rolling-band-free source coverage was 100% for all but 10 channels in the first 10 days of Q1 data.

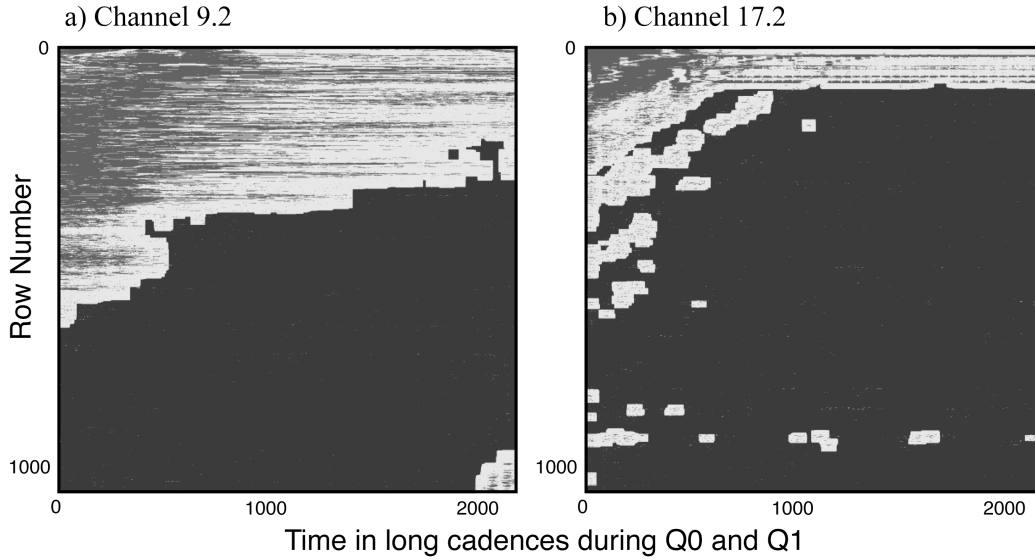


Figure 7. Example of rolling band flagging results from channels a) 9.2 and b) 17.2 for combined Q0-Q1. Varying shaded areas represent the degree beyond threshold level that the data indicates a rolling band, white regions are the padding around the offending regions and black areas are unflagged. The scales are image row vs. time as measured in long cadence periods.

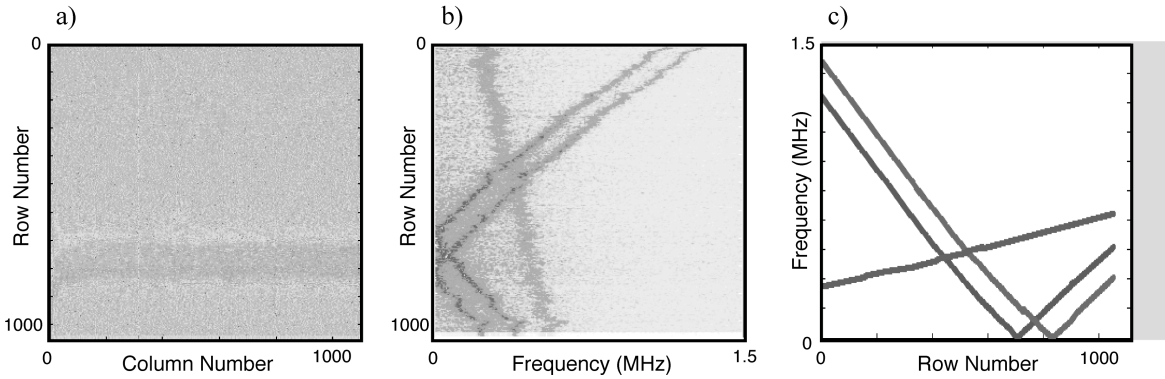


Figure 8. Example of moiré pattern detection algorithm. a) Difference between 2 full frame images from channel 17.2 during Q0. A rolling band is apparent in the row 700-850 range. b) Row-by-row FFT of image difference indicating 3 frequency traces. The 2 similar curves indicate a shift in frequency profile vs. row between the 2 FFIs. Localized frequency changes are the result of readout temperature response to bright sources (scene-dependent artifacts). c) The 3 frequency traces are detected by the filtering and detection algorithm, which is designed to avoid confusion when traces cross. Plots are derived from difference in 2 Channel 9.2 FFIs from Q2.

Effectiveness of the moiré pattern detection algorithm is illustrated in figure 8. The differenced FFI pair shown in figure 8.a shows clear moiré pattern features after subjection to a row-by-row FFT as shown in figure 8.b. The MPD detection and flagging algorithm identifies the distinct frequency traces in the images (figure 8.c) and determines the amplitudes at each frequency in each subwindow, as discussed in the algorithm description. Similar algorithms generate traces for the difference between the masked and virtual smear for each long cadence to enable the frequency matching and amplitude scaling between full images and long cadences. These scaled amplitudes are used in the threshold checking algorithm to flag data which is affected by moiré patterns.

When applied to all the readout channels, the algorithm detected peak-to-peak moiré amplitudes ranging from 0 to 1.2 DN/pixel/read over all possible moiré frequencies (0-1.5 MHz). The moiré-pattern-drift-free source coverage was 100% for 65 module-outputs and 0% for 6 module-outputs during the first 10 days of quarter-1. The source coverage in the remaining 13 readout channels varied from 17% to 86%.

## 5. CONCLUSIONS AND FORWARD WORK

The prototype algorithms described above successfully detect and correct or flag the pattern noise sources observed in the *Kepler* data stream. The following sections summarize the net impact of the image artifacts on the total *Kepler* field-of-view and outline forward plans for algorithm implementation, as well as mitigation of scene dependent artifacts.

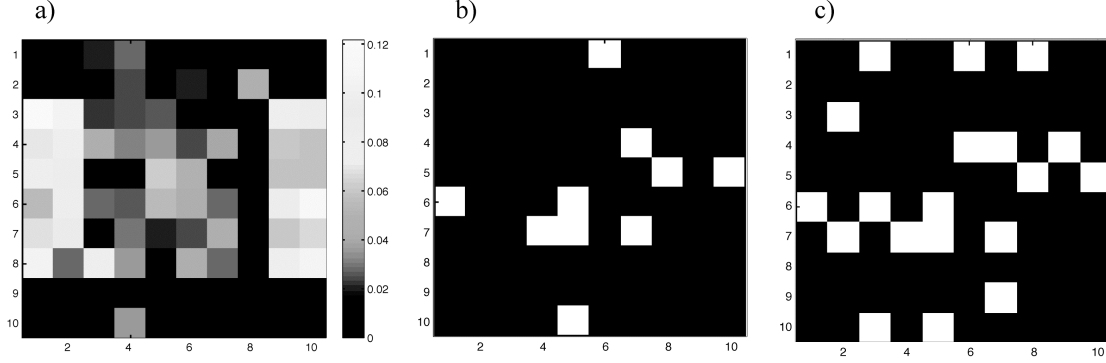


Figure 9. a) FGS cross-talk afflicted channels, with gray-level indicating the fraction of targets requiring corrections of 0.02 DN/pixel/read or greater. b) Channels exhibiting rolling bands are highlighted in white. c) Channels exhibiting moiré pattern drift are highlighted in white. Channel layout matches Figure 1a.

### 5.1 Summary

Results based on the first month of *Kepler* science data collection from all channels are consistent with prelaunch expectations. While 17.4% of targets' optimal apertures include modeled FGS clock intervals, only 3.7 % of targets require corrections of 0.02 DN pixel<sup>-1</sup> read<sup>-1</sup> or greater over the entire 34 day period, during which the temperature changed by 1°C. Rolling bands were evident on 10 channels during the first 10 days of this period, resulting in flagging of 3.0 % of the field-of-view. The rolling bands move with time, so the shorter interval gives a total flagged area representative of the affected size at any given instant. During the same interval moiré pattern afflicted 19 channels, resulting in flags over 12.7% of the field-of-view. This value is ~30% larger than prelaunch estimates and probably stems from the fact that in-flight temperatures are more stable, and therefore, the temperature profile during a single *Kepler* read is more repeatable. This would lead to greater coherence in the co-added frames, extending the moiré patterns further across the image rows. Figure 9 shows which *Kepler* channels are affected by the various pattern noise sources.

### 5.2 Implementation in the *Kepler* Science Processing Pipeline

Prior to launch, artifact mitigation was considered a top priority for *Kepler* pipeline implementation. This plan is undergoing a reprioritization based on in-flight observations of photometric performance and evaluation of the effects of other systematic errors and corrections. As a result, it is likely that these prototype algorithms will be delayed to a later pipeline development phase, assuming funding is available. Nevertheless, the completed prototype can be applied to any and all *Kepler* data to evaluate the impact of pattern noise on planet detection and measured parameter uncertainties.

### 5.3 Plans for scene dependent artifacts

The next step in artifact mitigation is to evaluate the impact of scene dependent artifacts on the field-of-view. Certain artifact features are temporarily distorted by the thermal transients introduced into the sensitive readout circuits by bright stars which saturate pixels. These stars produce a smear pattern extending over several rows in a given column. Specifically, the moiré pattern frequency changes abruptly and the amplitude increases sharply for many tens of pixels following these features. Pixels in the wake of these saturated segments are thereby subject to increased black-level variation. The affected regions are not directly detectable in the collateral regions and thus will require FFI images for detection. An algorithm which measures the Fourier peak amplitude in short spatial segments in these regions in differenced FFIs is expected to be effective at flagging the at-risk regions of MPD affected channels. In addition, variability of the bright stars changes the saturated segment's length producing a large dynamic range in the end pixels. Variation in these pixels produces variation in the downstream pixels in the same rows as a result of LDE undershoot. Since undershoot is only corrected in collected pixels, it would also be necessary to flag these areas based on FFI images and the known characteristics of bright variable stars in the field of view. In many cases it is possible to monitor the

variability of the bright stars in the smear collateral regions on a continuous basis. Channels typically have only ~500 saturated and near saturated pixels so even if these effects extend for 100 pixels downstream in the readout chain of each bright pixel, <5% of the field-of-view would be affected. Nevertheless, the need for automation in the detection of planet candidates in the >150000 *Kepler* targets supports the need to flag these at-risk regions.

## APPENDIX

This appendix defines the terms in the spatial model. We are modeling several discrete components with only the leading ( $C \leq 12$ ) and trailing black so it will be convenient to define the following discrete delta functions:

$$\delta_{XY} = \begin{cases} 1 & \text{if } X = Y \\ 0 & \text{if } X \neq Y \end{cases}$$

$$\delta_{C \in LB} = \begin{cases} 1 & \text{if } C \leq 12 \\ 0 & \text{if } C > 12 \end{cases}$$

$$\delta_{R \in MS} = \begin{cases} 1 & \text{if } R \leq 20 \\ 0 & \text{if } R > 20 \end{cases}$$

$$\delta_{R \in VS} = \begin{cases} 1 & \text{if } R \geq 1045 \\ 0 & \text{if } R < 1045 \end{cases}$$

$$\delta_{\{R,C\} \in US} = \begin{cases} 1 & \text{if } \{R,C\} \in \text{undershoot ARPs} \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_{\{R,C\} \in i}^F = \begin{cases} 1 & \text{if } \{R,C\} \in \text{pixels with FGS frame sequence number } i \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_{\{R,C\} \in i}^P = \begin{cases} 1 & \text{if } \{R,C\} \in \text{pixels with FGS parallel sequence number } i \\ 0 & \text{otherwise} \end{cases}$$

With these we can define the various modeled terms for LC fits:

$$f_{RC} = (C_1^f + \delta_{C \in LB} C_2^f) \ln \left( \frac{R}{R_{\log}} + 1 \right) + (C_3^f + \delta_{C \in LB} C_4^f) e^{-\frac{R}{R_{\exp}}}, \text{ where } R_{\log} \text{ and } R_{\exp} \text{ are constants}$$

$$g_C = \sum_{i \in C_{LB}} \delta_{Ci} C_i^g + \delta_{C \in LB} C_0^g, \text{ where } C_{LB} \text{ is the list of discretely modeled leading black columns}$$

$$W_{RC}^F = \sum_{i \in FGS-F} \delta_{\{R,C\} \in i}^F (C_{1,FGS-F}^{WF} + \delta_{C \in LB} C_{2,FGS-F}^{WF}), \text{ where FGS-F are modeled FGS frame clock states}$$

$$W_{RC}^P = \sum_{i \in FGS-P} \delta_{\{R,C\} \in i}^P (C_{1,FGS-P}^{WP} + \delta_{C \in LB} C_{2,FGS-P}^{WP}), \text{ where FGS-P are modeled FGS parallel clock states}$$

$$U_{RC} = \delta_{\{R,C\} \in US} (C_0^U + \sum_{j=C-N_{US}}^{C-1} C_{C-j}^U X_{Rj}), \text{ where } X_{Rj} \text{ is a measured pixel signal value, and } N_{US} \text{ is the}$$

number of modeled undershoot columns.

RCLCs are used to measure the column dependence of the black level using similar terms:

$$f_{RC} = U_{RC} = 0$$

$$g_{RC} = \sum_{i \in C_{LB}} \delta_{Ci} (C_{1i}^g + \delta_{R \in MS} C_{2i}^g + \delta_{R \in VS} C_{3i}^g) + \delta_{C \in LB} ((C_{10}^g + \delta_{R \in MS} C_{20}^g + \delta_{R \in VS} C_{30}^g) +$$

$$(C_{11}^g + \delta_{R \in MS} C_{21}^g + \delta_{R \in VS} C_{31}^g)C + (C_{12}^g + \delta_{R \in MS} C_{22}^g + \delta_{R \in VS} C_{32}^g)C^2),$$

where  $C_{LB}$  is the list of discretely modeled leading black columns

$$W_{RC}^F = \sum_{i \in FGS-F} \delta_{\{R,C\} \in i}^F (C_{1,FGS-F}^{WF} + \delta_{R \in MS} C_{2,FGS-F}^{WF} + \delta_{R \in VS} C_{3,FGS-F}^{WF}), \text{ where FGS-F are modeled FGS frame clock states}$$

$$W_{RC}^P = \sum_{i \in FGS-P} \delta_{\{R,C\} \in i}^P (C_{1,FGS-P}^{WP} + \delta_{R \in MS} C_{2,FGS-P}^{WP} + \delta_{R \in VS} C_{3,FGS-P}^{WP}), \text{ where FGS-P are modeled FGS parallel clock states}$$

Cadence-by-cadence fits of the masked-virtual smear differences employ this RCLC pixel model to monitor unexpected changes in the low-spatial-frequency column dependence of the black level and FGS crosstalk variation with column.



## REFERENCES

- [1] Borucki, W. J., et al., "Kepler Planet-Detection Mission: Introduction and First Results," *Science* 327(5968), 977- (2010).
- [2] Koch, D. G., et al., "Kepler Mission Design, Realized Photometric Performance, and Early Science ," *ApJ.* 713(2), L79-L86 (2010).
- [3] Argabright, V. S., et al., "The Kepler photometer focal plane array 2008," *Proc. SPIE*, 7010, 76 (2008).
- [4] Van Cleve, J., & Caldwell, D. A., [Kepler Instrument Handbook], KSCI 19033-001 (Moffett Field, CA: NASA Ames Research Center), <http://archive.stsci.edu/kepler/>, (2009).
- [5] Caldwell, D. A., et al., "Instrument Performance in Kepler's First Months," *ApJ.* 713(2), L92-L96 (2010).
- [6] Jenkins, J. M., et al., "Overview of the Kepler Science Processing Pipeline," *ApJ.* 713(2), L87-L91 (2010).
- [7] Haas, M. R., et al., "Kepler Science Operations," *ApJ.* 713(2), L115-L119 (2010).
- [8] Caldwell, D. A., et al., "Kepler Instrument Performance: An In-flight Update," *Proc. SPIE*, 7731, (2010).
- [9] Jenkins, J. M., et al., "Initial Characteristics of Kepler Long Cadence Data for Detecting Transiting Planets," *ApJ.* 713(2), L120-L125 (2010).